Behavioral graded activity following first-time lumbar disc surgery: 1-year results of a randomized clinical trial

SPINE Volume 28, Number 16, pp 1757–1765

Behavioral Graded Activity Following First-Time Lumbar Disc Surgery

1-Year Results of a Randomized Clinical Trial

Raymond W. J. G. Ostelo, PT, PhD,*† Henrica C. W. de Vet, PhD,‡
Johan W. S. Vlaeyen, PhD,§ Maria R. Kerckhoffs, PT, MSc,*†‖ Willem M. Berfelo, MD PhD,¶
Pieter M. J. C. Wolters, PT,‖ and Piet A. van den Brandt, PhD*

Study Design and Objectives. In a randomized clinical trial, the effectiveness of behavioral graded activity was assessed as compared to usual care provided by physiotherapists for patients after first-time lumbar disc surgery (n = 105).

Summary of Background Data. Little is known about the effectiveness of rehabilitation programs following lumbar disc surgery. Most programs focus on biomechanical aspects, whereas psychosocial factors are hardly addressed. The aim of the behavioral graded activity program, which is an operant treatment, is to alter psychosocial factors such as fear of movement and pain catastrophizing, which might subsequently lead to improved functional status and higher rates of recovery. Behavioral treatments for patients following lumbar disc surgery have not yet been assessed in a randomized clinical trial.

Methods. Inclusion criteria: age between 18 and 65 years; first-time lumbar disc surgery; restrictions in normal activities of daily living. Exclusion criteria: surgical complications and confirmed and relevant underlying diseases. Outcome assessment took place at 6 and 12 months after randomization.

Results. Six months after randomization, 62% of the patients had recovered following usual care *versus* 65% of the patients following behavioral graded activity. After 12 months, 73% and 75%, respectively, had recovered. Differences between intervention groups, 3% and 2% respectively, after 6 and 12 months are not statistically significant. Furthermore, there were no differences between the two groups regarding functional status, pain, pain catastrophizing, fear of movement, range of motion, general health, social functioning or return to work. After 1 year, 4 of the behavioral graded activity cases had undergone another operation *versus* 2 of usual care cases.

Conclusion. Both fear of movement and pain catastrophizing seem to be unaffected by either treatment in these patients. It is concluded that treatment principles derived from theories within the field of chronic low back pain might not apply to these patients. After 1 year of follow-up, there were no statistically significant or clinically relevant differences between the behavioral graded activity program and usual care as provided by physiotherapists for patients following first-time lumbar disc surgery. [Key words: rehabilitation, lumbar disc surgery, behavioral treatment, graded activity, randomized clinical trial] **Spine 2003;28:1757–1765**

Only a few good quality studies exist that address the nature and frequency of persisting or recurrent symptoms following lumbar disc surgery. The published figures, however, vary widely, ranging from 22% to 45% of patients reporting residual sciatica after lumbar disc surgery and 30% to 70% of patients reporting residual low back pain.[1–10] Persisting symptoms mainly consist of pain, motor deficits, and a decreased functional status. Rehabilitation following lumbar disc surgery is an important tool in order to minimize these complaints. Although various treatments have been suggested, little is known about the effectiveness of postsurgery rehabilitation. In a systematic review, 13 studies were identified that addressed the effectiveness of active rehabilitation following lumbar disc surgery.[11] These studies were highly heterogeneous with regard to the start of the rehabilitation program ranging from 2 days up to more than 12 months postsurgery. Also, duration and intensity of the interventions differed widely, from a 1-week postsurgery simple straight leg-raising regimen[12] to a 3-month rehabilitation program.[13] The results of the systematic review[11] show that there was no absolutely convincing evidence for the effectiveness of any treatment. However, based on two high quality randomized controlled trials (RCTs),[14,15] some evidence was identified for short-term effectiveness of intensive exercise programs that start 4 to 6 weeks postsurgery as compared with mild exercise programs. No relevant differences could, however, be demonstrated on long-term follow-up. It has been suggested that high intensity exercise programs confront patients with their fears and insecurities and that they may learn that symptoms related to training are not necessarily dangerous.[15]

All interventions in the included studies are focused on biomechanical aspects such as range of motion or muscle strength, whereas psychosocial factors (*e.g.*, pain catastrophizing or fear of movement) are hardly addressed. This is remarkable because most of the patients studied can be considered to be chronic pain patients suffering from symptoms for a period of several months before surgery. In these patients, psychosocial aspects

From the *Department of Epidemiology, Maastricht University, †Hogeschool Zuyd, Department of Physiotherapy, Heerlen, ‡EMGO Institute, VU University Medical Center Amsterdam, §Department of Clinical and Experimental Psychology, Maastricht University, and the Departments of ‖Physiotherapy and ¶Neurosurgery, University Hospital Maastricht, The Netherlands.
This research is supported by "Profileringsfonds" (PF-57; B96.1.996) of University Hospital Maastricht and "Stichting Annafonds Leiden."
Acknowledgment date: January 28, 2002. First revision date: July 19, 2002. Acceptance date: November 14, 2002.
Address correspondence to Raymond W. J. G. Ostelo, VU University Medical Center, Institute for Research in Extramural Medicine (EMGO), Van der Boechortstraat 7, Amsterdam 1081 BT, The Netherlands. E-mail: R.Ostelo.emgo@med.vu.nl

Table 1. Baseline Characteristics of Both Treatment Groups

Characteristics	Usual Care (n = 53)	BGA (n = 52)
Age (yrs)	43.7 (8.8)	42.8 (8.8)
No. (%) of women	19 (35.8)	26 (50.0)
Type of surgery: standard discectomy (%)	70	78
Level of surgery		
L4–L5 (%)	43.1	49.0
L5–S1 (%)	52.9	44.9
Surgical findings		
Sequester (%)	38.5	38.8
Protrusion (%)	51.9	49.0
Duration hospitalization (days)	7.7 (2.7)	7.0 (1.3)
Duration of symptoms before surgery		
≤3 mos: [no. (%)]	9 (17)	7 (13)
>3 mos: [no. (%)]	44 (83)	45 (87)
Physiotherapy immediately after surgery (%)	33	42
Taking pain medication [no. (%)]	18 (34)	20 (38)
Physiotherapy before surgery [no. (%)]	30 (56.6)	32 (60.4)
Other treatment before surgery [no. (%)]	16 (30.2)	13 (23.9)
Paid employment [no. (%)]	37 (69.8)	47 (90.4)
Confidence with regard to recovery in general [no. (%)]		
Great deal	22 (41.5)	21 (40.4)
Moderate amount	18 (34.0)	24 (46.2)
No	3 (5.7)	3 (5.8)
Don't know	10 (18.9)	4 (7.7)
Expectancy of allocated treatment (0–10 points)*	6.9 (1.6)	6.9 (1.1)
Negative affectivity (0–14 points)	3.7 (4.0)	4.2 (4.1)
Outcome measures		
Roland Disability Questionnaire (0–24 points)	13.5 (4.5)	14.5 (3.7)
Tampa Scale (17–68 points)	36.9 (6.8)	35.9 (6.3)
Pain Catastrophizing Scale (0–52 points)	16.9 (11.7)	17.1 (10.2)
Main Complaint (0–100)	67.4 (15.4)	71.1 (16.5)
Pain in back (No. (%))	45 (85)	48 (92)
Severity of pain in back (0–100)	46.7 (27.3)	43.4 (30.0)
Sciatica [no. (%)]	52 (98)	50 (96)
Severity of sciatica (0–100)	41.3 (30.8)	39.0 (28.2)
Range of motion	81.2 (22.7)	78.1 (22.6)
General Health (subscale SF-36)	65.6 (20.0)	68.2 (18.4)
Social Functioning (subscale SF-36)	59.4 (25.3)	56.7 (26.6)

Values are means with standard deviations between brackets unless stated otherwise.
* Administered after 2 treatments.
BGA = behavioral graded activity.

such as pain catastrophizing or fear of movement may be very important.[16] Directing interventions towards these psychosocial aspects might, therefore, be promising.

Based on recent studies,[16,17] it was hypothesized that behavioral graded activity (BGA), which is an operant treatment, may alter fear of movement and pain catastrophizing, and thus may subsequently lead to improved functional status and higher rates of recovery. So far, however, this has only been studied for patients with nonspecific chronic low back pain. As far as we know, there is no randomized trial that evaluates a behavioral program following lumbar disc surgery. Therefore, the aim of this randomized clinical trial is to evaluate whether a BGA program is more effective than usual care (UC) as provided by physiotherapists when applied to patients following first-time lumbar disc surgery. The posttreatment measurement, 3 months after randomization, revealed that on global perceived effect the BGA program performed statistically significantly worse than UC (19.3% difference, 95% confidence interval [CI]: 0.1%; 38.5%). For all other outcome measures (*e.g.*, functional status or pain), there were no statistically significant or clinically relevant differences between the groups.[18] This paper reports the outcomes at 6 and 12 months of follow-up.

■ Materials and Methods

Study Design and Randomization. In an RCT, the effect of BGA and UC provided by physiotherapists in patients following first-time lumbar disc surgery was compared on functional status and recovery rates. The Medical Ethics Committee of the University Hospital Maastricht approved the study protocol. An extensive description of the design, background to the intervention, and outcome measures has been published elsewhere.[19] By using opaque, sealed, and coded randomization envelopes, which were based on computer-generated randomization lists, the research assistant who also performed the outcome assessments (M.R.K.) was blinded. To assess the success of randomization, several important prognostic factors (Table 1) were measured at baseline including scores for all outcome measures. Outcome measurements took place posttreatment (3 months after randomization) to detect short-term effects and 6 and 12 months after randomization to detect long-term effects. In this paper, we present the results of 6 and 12 months of follow-up.

Selection of Patients and Informed Consent. Indication for surgery was persisting symptoms (pain and/or sciatica) for 4 to 6 weeks, consistent with clinical findings and findings on computed tomography (CT) or magnetic resonance imaging (MRI). Patients were scheduled for a routine postsurgery visit to the neurosurgeon 6 weeks after surgery. In case of persisting symptoms (severe leg or back pain, motor deficits, restriction of activities of daily living [ADL], and/or work), they were referred for physiotherapy, the neurosurgeon checked the eligibility criteria, and patients received oral and written information about the study. The research assistant then provided further details about the study and re-evaluated their eligibility. Inclusion criteria were: age between 18 and 65 years; first-time lumbar disc surgery (one level only); and complaints (*e.g.*, pain) restricting their ADL and/or work. Patients were excluded in the case of surgical complications, to be judged by the neurosurgeon and based on preset criteria (nerve root lesion, loss of cerebrospinal fluid, loss of more than 600 mL blood), in case of confirmed and relevant underlying diseases that influenced ADL (*e.g.*, stenosis, malignancies, M. Bechterew, M. Scheuerman), or if one of the treatments was contraindicated (*e.g.*, due to respiratory complaints). If patients were eligible and willing to participate, informed consent was signed.

Interventions. Both interventions have been described extensively elsewhere.[19] Behavioral graded activity is an operant therapy using graded activity and positive reinforcement in order to increase healthy behavior and to decrease pain behavior.[20,21] It is based on time-contingency management as described in more detail by Fordyce,[20] Fordyce *et al*,[22] and applied by Lindström *et al*.[23] The term "behavioral graded activity" for this program emphasizes the behavioral component rather than merely physical training principles. Physiotherapists who attended a 2-day practical training course and 2 refreshment meetings during the study provided the treatment. The essence of BGA was to establish individually graded exercise trainings, based on baseline measurements performed at intake, and let patients experience that it is safe to increase activity levels. During initial baseline measurements, patients were asked to perform activities (selected by the patients themselves) or exercises until reaching (pain) tolerance. Then, patients set their own individual treatment goals for these activities or exercises. The next step was to set quotas (time contingent), which were systematically increased towards the preset goal. Quotas were set by the patient, supervised by the physiotherapist, and were not to be overperformed nor underperformed. First quotas were slightly under baseline level, to assure that patients' initial experiences, while performing exercises, were successful. This enhances motivation and enables positive reinforcement, which is one of the key principles in the operant conditioning theory. In this way, a patient-tailored, individual BGA program was developed. Patients had to practice at home and document activities or exercises on performance charts that were discussed at each treatment session.

The content of the other treatment condition was determined after extensive interviews and two consensus meetings with the participating physiotherapists. Therefore, this reflects the care as usual among physiotherapists in treating patients following a lumbar disc surgery, and we labeled this treatment as usual care. In general, the whole spectrum of techniques used by the physiotherapists was included, which, in our opinion, is sensible when studying UC. The main topics of the UC could be outlined from the treatment registration forms: all physiotherapists instructed their patients to exercise trunk muscles to increase strength and stability. The exercises aimed, furthermore, at increasing the levels of ADL. Sixty-five percent of the physiotherapists explicitly instructed patients how to lift, sit, and stand and how to perform other kinds of ADL. Forty-five percent of the physiotherapists used some kind of electrotherapy in at least three (or more) sessions to decrease pain and muscle tone. Thirty percent of the physiotherapists used some kind of hands-on technique (massage or manipulations) in some of the treatment sessions (ranging from 2 up to 13 sessions) to decrease pain and muscle tonus.

The physiotherapists documented every session on treatment registration forms in both treatment conditions. These treatment conditions consisted of a maximum of 18 sessions (30 minutes each) within a period of 3 months. Contrary to the BGA, UC physiotherapists were allowed to stop treatment as soon as the complaints had disappeared and the treatment goals had been achieved, thus complying with usual care principles.

Integrity Check of the Interventions. The following features were used to define the difference between BGA and UC. First of all, BGA is based on systematic baseline measurements, whereas UC relies on anamnesis and physical examination. Secondly, BGA management is time contingent once quotas have been set, whereas UC evaluates reactions on previous treatments and eventually adapts treatment intensity based on this pain contingent evaluation. Thirdly, BGA relies on specific behavioral components: goal-setting by patients, performance charts, systematic appraisal and reinforcement of health behaviors, and extinction of pain behavior. To evaluate the differences between both treatment arms, audio tapes were recorded and rated by three blinded experts. Three prerecorded sound samples, in our opinion containing an optimal BGA treatment, were also included in order to evaluate the scoring system.

Prognostic Factors and Outcome Measurements. Demographic and clinical information (Table 1) was retrieved from the patient files. At baseline, the duration of complaints, medication, previous treatments, and professional occupation were documented. At baseline, patients were also asked to rank their level of confidence with respect to recovery (great deal, moderate amount, no, don't know). After 2 treatment sessions, patients were asked to what extent they expected the allocated treatment to be beneficial to them (10-point Likert scale: 0 = expects no benefit at all, 10 = absolutely convinced of benefit).[24] Negative affectivity was assessed with the Negative Emotionality (NEM) subscale (14 items, 2-point scale) of the Multidimensional Personality Questionnaire.[25,26] Negative affectivity correlates with psychosomatic symptoms, anxiety, worries, and poor role adjustment. High NEM scores denote high levels of negative affect.

Outcome Measurements. Primary outcome measures were: 1) global perceived effect (GPE) rated on a 7-point scale (1 = completely recovered, 7 = worse than ever) to assess recovery. These ratings were dichotomized into "improved" ("completely recovered" and "much improved") *versus* "not improved" ("slightly improved," "not changed," "slightly worsened," "much worsened," "worse than ever"). *A priori* a 20% difference between groups in dichotomized "improvement" rates was considered clinically relevant[19]; and 2) the Roland Disabil-

ity Questionnaire (RDQ)[27] measures low back specific functional status, and the Dutch translation was validated.[28,29]

Secondary outcome measures were: 1) fear of movement (Tampa Scale for Kinesiophobia [TSK]).[30] The Dutch translation has a fair and consistent internal validity.[31] 2) The Pain Catastrophizing Scale (PCS)[32] that measures pain catastrophizing (viewing pain as extremely threatening). Intensity of low back pain or sciatica was scored on a visual analogue scale (VAS). Relevance, validity, and reliability of the VAS are commonly accepted in the area of low back pain.[33–35] At baseline, patients selected two important ADL activities that were severely hampered by their symptoms, in a standardized fashion. These were called Main Complaints (MC).[36] The severity was scored on a VAS. General health and social functioning were evaluated by using the corresponding subscales of the SF-36.[37] The Dutch translation by van der Zee *et al*[38] showed satisfactory validity and reproducibility. Range of motion (ROM) (flexion, extension) of the lumbar spine is measured by the Cybex Electronic Digital Inclinometer (EDI-320) that proved to be acceptably reproducible, especially for flexion.[39–41] Occurrences of reoperations, use of medication, and health care utilization were recorded at the 1-year follow-up. Finally, at the 6- and 12-month follow-ups, all subjects that reported to be employed at baseline were asked whether they had returned to work.

Analysis. Treatment registration forms were screened for protocol violations. In the BGA program, protocol violations were defined as: use of passive treatment methods, not fulfilling quotas more often than twice, and cointerventions by other health care providers (*e.g.*, neurosurgeons). In UC, only major interventions such as ceasing treatment according to the advice of the neurosurgeon were recorded as protocol violations. Statistical analyses were carried out according to the intention-to-treat principle: all patients, including dropouts, remained in the group to which they were assigned by randomization. The cause of dropping out determined the replacement procedure: 1) patients were assigned the mean value of their group if there was no association with allocated treatment (*e.g.*, patients moved out of the catchment area); 2) patients received negative scores if they had more pain, or in case of suspicion of a (new) herniated disc; 3) patients received positive scores if, for example, they had returned to work full-time. For substitution of negative or positive scores, we used the 10th or 90th percentile scores. If 2 out of the 3 blinded and independent experts attributed the same substitution value, then this value was used. A per-protocol analysis was performed which was restricted to those patients who complied with the treatment protocol. For all analyses, SPSS 9.0 for Windows (SPSS Inc., Chicago, IL) was used. For outcome measures collected at baseline, the differences between the follow-up measurements and the baseline score were calculated for each individual and these change scores were compared between the two groups, using the Student *t* test for statistical significance. For outcome measures without baseline measurement, the differences between groups at follow-up were analyzed. Group differences and two-tailed 95% CI were calculated for all outcome measures. In order to adjust for possible baseline differences, a multiple linear regression analysis for continuous outcome measures was performed. The proportion of subjects who had returned to work at the 6- and 12-month follow-ups was analyzed using the χ^2 test, and α was set at 0.05. Subgroup analyses were carried out to determine whether particular categories of subgroups showed different outcomes than the treatment groups as a total. Dichotomized subgroups were formed according to negative affectivity (cutoff on NEM: 7), fear of movement (cutoff on TSK: 40), and confidence of patients about their recovery ("great deal" *vs.* "moderate amount," "no," and "don't know"). With respect to the audio tapes for assessing the treatment integrity of both interventions, first agreement between the three experts was calculated on the original VAS scores by means of the Pearson r. Then, for each characteristic, the percentage correctly classified was calculated.

■ Results

From November 1997 until December 1999, 671 patients were screened for eligibility in the 4 participating hospitals in the south of the Netherlands. More than half of the patients recovered after first-time lumbar disc surgery and suffered no further substantial symptoms (57%); 141 patients (21%) were excluded because of various reasons: patients, although operated on in a participating hospital, were not living in the catchment area (n = 40), not motivated to participate (n = 32), too old (n = 30), presented with comorbidities (n = 22), language problems (n = 8), previous lumbar surgeries (n = 7), and due to insurance problems (n = 2). We failed to trace 43 patients (6%). In total, 105 (16%) patients were eligible and signed informed consent. Figure 1 summarizes the patient flow through the study.

On average, 15.5 (2.1) treatment sessions were realized in UC *versus* 14.8 (2.6) in BGA. On the posttreatment measurement, eight patients dropped out: one from the UC group and seven from BGA. The UC patient disappeared after two treatments without stating any reason and was therefore assigned the mean values of the UC group. The two BGA patients withdrew from the study due to aggravated symptoms; negative scores therefore substituted their values. One BGA patient showed an exacerbation of symptoms before treatment and another BGA patient suffered from rheumatic symptoms, a disease he had not mentioned before randomization. These reasons were considered not to be related to the postoperative treatment; therefore, they were assigned mean values. One BGA patient reported to be completely pain-free after two treatment sessions and was no longer motivated to continue the study, and one patient stepped out when resuming his professional occupation full time (without residual signs or symptoms) due to lack of time and motivation. One BGA patient withdrew because of personal circumstances and had actually recovered after five treatment sessions. The values of these three patients were substituted by positive values. After 6 months follow-up, 2 more BGA-patients dropped out: one underwent an operation for an intestinal disorder, and the other patient dropped out without obvious reasons and did not react to several voice mail requests. Both patients were assigned mean values. Another UC patient dropped out because of aggravated symptoms; negative values were used for substitution. After 12 months of follow-up, another BGA patient

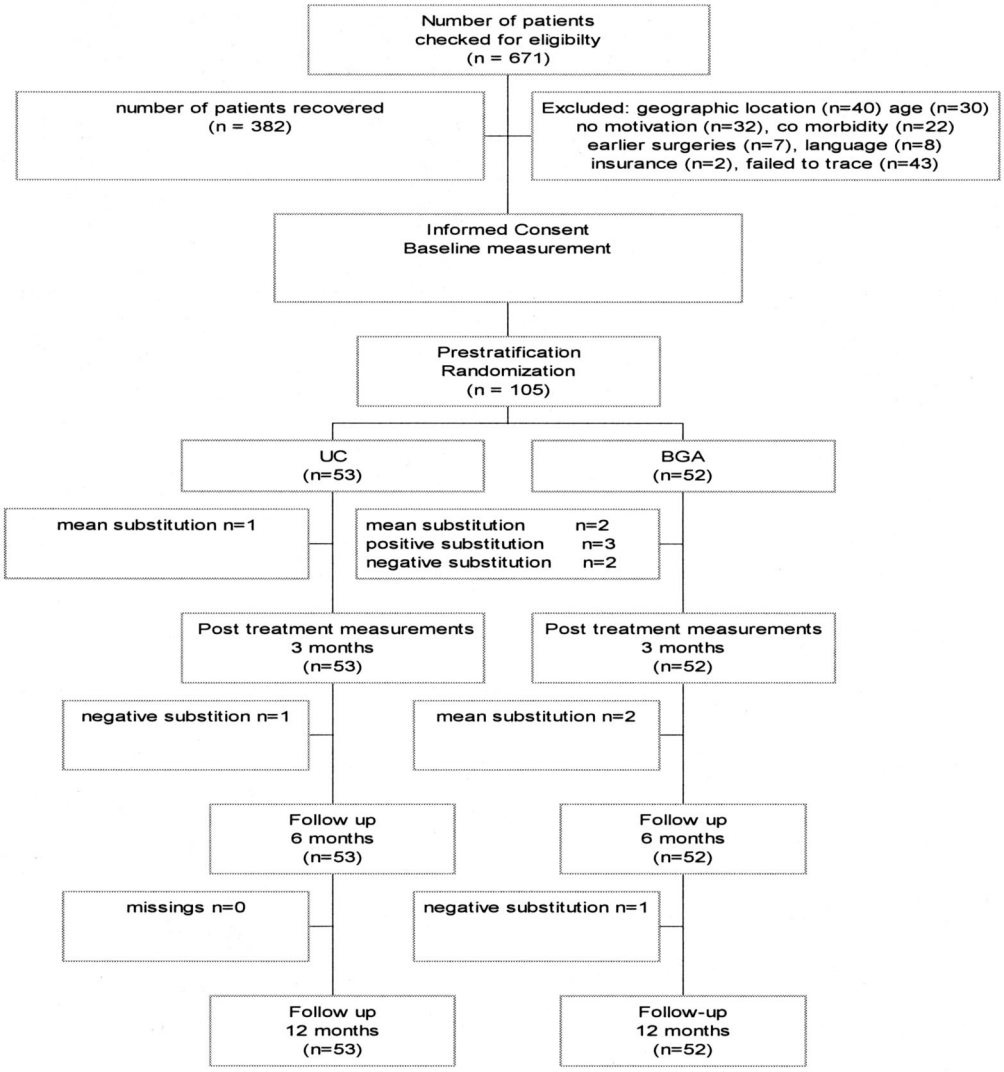

Figure 1. Patient flow through study.

dropped out due to aggravated symptoms; negative values were used for substitution.

Distributions of baseline characteristics of both groups are presented in Table 1. Randomization yielded two prognostic comparable treatment groups at baseline. At the posttreatment measurements, it appeared that in each group, in one case, surgery was repeated at the same level as initially.[18] After 6 months of follow-up, 2 more operations were repeated, again one in each group. After 12 months of follow-up, another 2 operations were repeated in the BGA group. The total number of repeated operations at 12 months of follow-up was 4 in BGA and 2 in UC.

Table 2 presents the results of the effectiveness of the interventions. After 6 months of follow-up, 62% of the patients had recovered in UC *versus* 65% in BGA. After 12 months of follow-up, these rates were 73% in UC and 75% in BGA. The differences between interventions, 3% and 2% respectively, after 6 and 12 months of follow-up are neither statistically significant nor clinically relevant. Although the improvements on the disability scores, as measured by the RDQ, are statistically significantly different within the groups and clinically relevant, differences between groups are not. After 12 months of follow-up, a statistically significant difference occurs in favor of the BGA on pain catastrophizing, although the 3.3-point difference (95% CI: −6.4; −0.1) on this scale bears no clinical importance.[42] The scores on the TSK revealed no statistically significant or clinically relevant differences between the groups. Furthermore, Table 2 shows that there is neither a statistically significantly nor a clinically relevant difference between the two groups on any of the other outcome measures.

Adjustments for *a priori* identified covariables (baseline scores of RDQ, TSK, NEM, pain back, PCS, duration of complaints, and level of confidence in recovery) did not alter the results substantially. Therefore, we present only the unadjusted results. The per-protocol analyses were restricted to 78 patients: 45 patients in UC and 33 patients in BGA. The prognostic comparability between the intervention groups that qualified for the per-protocol analyses was quite similar to the results as

Table 2. Results at 6 and 12 Months of Follow-up

	6 Months Follow-up			12 Months Follow-up		
	Improvement Within Groups			Improvement Within Groups		
Outcome Measures	UC (n = 53)	BGA (n = 52)	Mean Difference BGA − UC [95% CI]	UC (n = 53)	BGA (n = 52)	Mean Difference BGA − UC [95% CI]
Global Perceived Effect (%)*	62%	65%	3.0% [−16.3; 22.4]	73%	75%	2.0% [−19.5; 15.7]
Roland Disability Questionnaire (0–24)†	−6.1 (5.6)	−6.4 (5.8)	−0.3 [−2.5; 2.0]	−7.0 (5.3)	−7.0 (5.5)	0.0 [−2.1; 2.1]
Pain Catastrophizing Scale (0–52)†	−4.4 (9.5)	−4.7 (7.4)	−0.3 [−3.6; 3.0]	−4.0 (7.9)	−7.3 (7.7)	−3.3 [−6.3; −0.2]‖
Tampa Scale Kinisiophobia (17–68)†	−1.7 (6.6)	−2.8 (6.4)	−1.1 [−3.6; 1.4]	−2.6 (6.2)	−2.7 (6.5)	0.0 [−2.5; 2.4]
Main Complaint (0–100)†	−38.6 (25.8)	−44.2 (29.1)	−5.6 [−16.3; 5.0]	−44.3 (30.6)	−50.3 (27.4)	−6.0 [−17.3; 5.5]
Pain in back (0–100)†	−20.9 (31.6)	−13.7 (31.4)	7.2 [−4.9; 19.4]	−22.4 (33.0)	−17.6 (32.5)	4.8 [−7.9; 17.5]
Sciatica (0–100)†	−16.9 (27.6)	−14.0 (26.1)	2.9 [−7.5; 13.3]	−19.1 (34.0)	−16.9 (28.4)	2.3 [−9.8; 14.4]
Range of motion	17.4 (19.0)	13.3 (24.4)	−4.1 [−12.9; 5.0]	18.9 (21.5)§	20.1 (22.7)§	1.2 [−7.5; 10.0]
General Health‡	4.8 (15.0)	1.0 (15.0)	−3.8 [−9.7; 2.0]	5.2 (16.6)	3.0 (16.4)	−2.2 [−8.6; 4.2]
Social Functioning‡	23.1 (24.0)	18.8 (27.2)	−4.3 [−14.2; 5.6]	24.1 (25.1)	28.4 (27.7)	4.3 [−14.5; 5.9]

All outcome measures are presented in means with standard deviations between brackets, unless stated otherwise.
* Dichotomized global perceived effect.
† Negative values denote positive results for patients.
‡ Subscale SF-36.
§ 4 missing due to technical problems.
‖ Statistically significant differences between groups.
UC = usual care; BGA = behavioral graded activity; CI = confidence interval.

summarized in Table 1. In general, the per-protocol analyses resulted in slightly larger improvements within groups, but the between-group differences did not change substantially.

Subgroup analyses revealed no differences after 6 or 12 months of follow-up in outcomes for negative affectivity (cutoff on NEM: 7), fear of movement (cutoff on TSK: 40), or confidence of patients concerning their recovery ("great deal" *vs.* "moderate amount," "no," and "don't know").

After 1 year of follow-up, the number of patients using medication and health care utilization was recorded over the preceding 3 months. In UC, 31% used analgesics *versus* 21% in BGA. Over this same period, 33% in UC visited a general practitioner (GP) *versus* 30% in BGA. Also, the number of patients visiting a specialist (neurosurgeon, neurologist, or orthopedic surgeon) and a physiotherapist revealed no statistically significant or clinically relevant differences: 13% in UC *versus* 15% in BGA for specialists and 12% in UC *versus* 10% in BGA for physiotherapists. None of the abovementioned differences are statistically significantly different.

At the 6-month follow-up, 84.2% and 68.3% ($P = 0.10$), respectively, in UC and BGA had returned to work. At the 12-month follow-up, this proportion was 90% in the UC group *versus* 73.7% in the BGA group ($P = 0.61$).

Integrity Check of the Interventions

In general, the agreement (Pearson r) between experts was on average 0.65, ranging from 0.55 to 0.82 for the various characteristics, which was considered acceptable. Furthermore, the three prerecorded samples were scored in all instances as expected, thus supporting the validity of the scoring system. On average, 70% to 80% of the sound samples were classified correctly, meaning that 20% to 30% was classified wrongly (*e.g.*, UC samples were scored as if they were a BGA sample).

■ Discussion

Main Findings

The effectiveness of a BGA program as compared to UC was assessed in a randomized clinical trial (n = 105). The posttreatment measurement, 3 months after randomization, revealed that on GPE, the BGA program performed statistically significantly worse than UC (19.3% difference, 95% CI: 0.1%; 38.5%); for all other outcome measures, there were no relevant differences between the groups.[18] After 6 and 12 months, the differences between groups were negligible for all outcome measures. The statistically significantly difference on 3 months on GPE might be a change finding due to multiple testing. In BGA, four operations were repeated *versus* two in UC. There were no substantial differences with regard to medication use or in the number of patients visiting a health care provider after 1 year of follow-up. Although the proportion of subjects that had returned to work was higher in the UC group at the 6- and 12-month follow-up measurements, these differences were also not statistically significant.

Can Postsurgery Patients be Considered as Chronic Pain Patients?

A priori it was hypothesized that BGA would alter fear of movement and pain catastrophizing, which would subsequently lead to an improved functional status and higher rates of recovery in patients following first-time lumbar disc surgery. This assumption was primarily based on studies concerning fear of movement and pain catastrophizing in patients with chronic low back pain.[16] However, the results of the current study do not support this hypothesis. Both fear of movement and pain cata-

strophizing seem to be unaffected by either treatment in this population. One explanation might be that patients who underwent lumbar disc surgery are different from patients with chronic low back pain. For example, RDQ scores at baseline in the current study are relatively high compared to recent studies that include patients with chronic low back pain,[43,44] indicating a more severe degree of disability. Furthermore, although the majority OF patients in the current study had suffered symptoms for more than 3 months (83% and 87%, respectively, in UC and BGA) and are, therefore, usually labeled as chronic, the period after the operation might be more important to classify these patients. It could be argued that patients who had a new start in their low back pain episode, because they had undergone an operation but had not yet recovered after 6 weeks, might be considered as balancing on the threshold of becoming patients with chronic low back pain[45] and that treatment principles derived from theories within the field of chronic low back pain might therefore not apply to this group.

The results show that, although not statistically significant, the proportion of the return in the UC group was higher as compared to BGA. But some remarks have to be made. Firstly, return to work is possibly also an outcome that results from the surgery that patients underwent 6 weeks before inclusion of our trial. The natural course of recovery after surgery has probably influenced return to work to a great extent. Secondly, return to work was not a specific treatment goal of either BGA or UC. If patients are absent from work following lumbar disc surgery and if return to work is of paramount importance to them, more specific reintegration approaches are called for. Therefore, the results with regard to return to work have to be interpreted cautiously.

Methodologic Issues Concerning the Current Study

Missing values are always a considerable nuisance when analyzing results.[46] Unfortunately, no clear-cut solutions are available. In this study, 12 patients were lost to follow-up after 1 year. The reasons for not attending the follow-up measurements were taken into account because this procedure best anticipates the information at our disposal. In BGA, more protocol violations were recorded, as expected, because the definition of protocol violation in BGA was stricter compared to UC, where only major interventions were considered to be protocol violations. In general, the per-protocol analyses resulted in slightly more improvements within groups, but the between-group differences did not change substantially. One explanation might be that especially patients who do not show any improvement during treatment are more prone to violate the treatment protocol. Another reason might be that the treatment induces larger improvements, if compliance is good. At any rate, the between-group differences did not change substantially, and, therefore, we concluded that our findings are not influenced by protocol violations.

Treatment integrity of both interventions is an important issue in the current study. Despite the 2-day training course and refreshment meetings during the trial, BGA might still not have been delivered as planned. To change the behavior of health care providers might be as difficult as changing the behavior of patients. The analysis of audio tapes showed that the three experts classified 70% to 80% of the sound samples in the correctly. Although there was an overlap of 20% to 30% resulting in less contrast than we thought in advance, we do not believe that this overlap concealed any possible effect of BGA, as BGA did not show the slightest sign of being more effective than UC.

A no-treatment control group was not included because the aim of this study was to investigate whether BGA was more effective than UC as provided by physiotherapists, which is a standard prescription in the participating hospitals. Therefore, all patients who still suffered complaints at the 6-week routine postsurgery visit to the neurosurgeon were treated. Furthermore, it was considered inappropriate to withhold treatment from patients if they still had symptoms 6 weeks after surgery. But now, as the results show that there are no differences between both interventions, it is difficult to attribute improvements to either treatment. These results may represent the natural course after first-time lumbar disc surgery.

Comparison With Other Studies

Comparing the results of the current study with the systematic review that assessed the effectiveness of active rehabilitation programs following lumbar disc surgery[11] is difficult because none of the included studies incorporated behavioral treatment. However, in the current study, as well as in the systematic review, there are no relevant differences on long-term follow-up when comparing two active rehabilitation programs. Although active rehabilitation might be effective after a lumbar disc surgery, it is not yet clear which components should be present in such a program. There is not much evidence with regard to the optimal starting point for rehabilitation after lumbar disc surgery.[11] Should all patients be treated immediately after surgery, or is it more efficient and effective to wait for 4 to 6 weeks and then include only patients who have not yet recovered? In the current study, patients were included if they still suffered symptoms at the 6-week consultation, because this is the policy in the participating hospitals. Therefore, based on this study, it is not possible to make recommendations with regard to the optimal starting point for postsurgery interventions.

■ Conclusion

Both fear of movement and pain catastrophizing seem to be unaffected by either treatment in these patients. It is concluded that treatment principles derived from theories within the field of chronic low back pain might not apply to these patients. After 1 year of follow-up, there

were no statistically significant or clinically relevant differences between the BGA program and UC as provided by physiotherapists for patients following first-time lumbar disc surgery.

■ Key Points

- Little is known about the effectiveness of rehabilitation programs following lumbar disc surgery, and no randomized controlled trials exist that assess a behavioral treatment including patients following lumbar disc surgery.
- In a randomized clinical trial that included 105 patients following first-time lumbar disc surgery, the effectiveness of a behavioral graded activity program was assessed.
- After 1 year of follow-up, there were no statistically significant or clinically relevant differences between the behavioral graded activity program and usual care as provided by physiotherapists for patients following first-time lumbar disc surgery.

Acknowledgments

The authors thank the "Profileringsfonds" (Grant no. (PF-57; B96.1.996) of the University Hospital Maastricht and the "Stichting Annafonds" for their financial support. They also thank all participating neurosurgeons from the University Hospital Maastricht, Atrium Hospital Heerlen, Maasland Hospital Sittard, and Laurentius Hospital Roermond for referring patients to our study. The authors are grateful to all physiotherapists and patients for participating in this study and indebted to Rob Pelt and Jan Thomassen for training and monitoring the physiotherapists. They appreciate the assistance of Cobie Martens with logistic and data management.

References

1. Barrios C, Ahmed M, Arrotegui JI, et al. Clinical factors predicting outcome after surgery for herniated lumbar disc: an epidemiological multivariate analysis. *J Spinal Disord* 1990;3:205–9.
2. Dvorak J, Valach L, Fuhrimann P, et al. The outcome of surgery for lumbar disc herniation. II. a 4–17 year follow-up with emphasis on psychosocial aspects. *Spine* 1988;13:1423–7.
3. Hurme M, Alaranta H. Factors predicting the result of surgery for lumbar intervertebral disc herniation. *Spine* 1987;12:933–8.
4. Korres DS, Loupassis G, Stamos K. Results of lumbar discectomy: a study using 15 different evaluation methods. *Eur Spine J* 1992;1:20–4.
5. Manniche C, Asmussen KH, Vinterberg H, et al. Analysis of preoperative prognostic factors in first-time surgery for lumbar disc herniation, including Finneson's and modified Spengler's score systems. *Dan Med Bull* 1994;41: 110–5.
6. Manniche C, Asmussen KH, Vinterberg H, et al. Back pain, sciatica and disability following first-time conventional haemilaminectomy for lumbar disc herniation. Use of "Low Back Pain Rating Scale" as a postal questionnaire. *Dan Med Bull* 1994;41:103–6.
7. Pappas CTE, Harrington T, Sonntag VKH. Outcome analysis in 654 surgically treated lumbar disc herniations. *Neurosurgery* 1992;30:862–6.
8. Smulders WLM, van Loon G. Nabehandeling na lumbale discectomie of chemonucleolyse. Een 'out-patient' enquête. *Fysiopraxis* 1993;2:4–6.
9. Spangfort EV. The lumbar disc herniation. A computer-aided analysis of 2,504 operations. *Acta Orthop Scand Suppl* 1972;95:1421–95.
10. Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. *Spine* 1983;8:131–40.
11. Ostelo RW, de Vet HCW, Waddell G, et al. Rehabilitation after first time disk surgery. A systematic review within the Cochrane framework. *Spine* 2003;28:209–18.
12. Kitteringham C. The effect of straight leg raise exercises after lumbar decompression surgery—a pilot study. *Physiotherapy* 1996;82:115–23.
13. Alaranta H, Hurme M, Einola S, et al. Rehabilitation after surgery for lumbar disc herniation: results of a randomized clinical trial. *Int J Rehabil Res* 1986;9:247–57.
14. Danielsen JM, Johnsen R, Kibsgaard SK, et al. Early aggressive exercise for postoperative rehabilitation after discectomy. *Spine* 2000;25:1201–6.
15. Manniche C, Skall HF, Braendholt L, et al. Clinical trial of postoperative dynamic back exercises after first lumbar discectomy. *Spine* 1993;18:92–7.
16. Vlaeyen JW, Crombez G. Fear of movement/(re)injury, avoidance and pain disability in chronic low back pain patients. *Man Ther* 1999;4:187–95.
17. van Tulder MW, Ostelo RW, Vlaeyen JW, et al. Behavioral treatment for chronic low back pain: a systematic review within the framework of the Cochrane Back Review Group. *Spine* 2001;26:270–81.
18. Ostelo RW, de Vet HCW, Vlaeyen JWS, et al. Effectiveness of behavioral graded activity after first-time disc surgery. *Eur Spine J*, 2003, in press.
19. Ostelo RW, Köke AJ, Beurskens AJ, et al. Behavioral-Graded activity compared with usual care after first-time disk surgery: Considerations of the design of a randomized clinical trial. *J Manipulative Physiol Ther* 2000;23: 312–9.
20. Fordyce WE. *Behavioral Methods for Chronic Pain and Illnesses*. St. Louis, MO: Mosby; 1976.
21. Vlaeyen JW, Haazen IW, Schuerman JA, et al. Behavioral rehabilitation of chronic low back pain: comparison of an operant treatment, an operant-cognitive treatment and an operant-respondent treatment. *Br J Clin Psychol* 1995;34:95–118.
22. Fordyce WE, Fowler RS Jr, Lehmann JF, et al. Operant conditioning in the treatment of chronic pain. *Arch Phys Med Rehabil* 1973;54:399–408.
23. Lindström I, Öhlund C, Eek C, et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study an operant-conditioning behavioral approach. *Phys Ther* 1992;72:279–93.
24. Vlaeyen JW, Kole-Snijders AM, van Eek H. *Chronische Pijn en Revalidatie*. Hoensbroek: Bohn Stafleu Van Loghum; 1996.
25. Stegen K, Neujens A, Crombez G, et al. Negative affect, respiratory reactivity, and somatic complaints in a CO2 enriched air inhalation paradigm. *Biol Psychol* 1998;49:109–22.
26. Tellegen A. *Brief Manual for the Multidimensional Personality Questionnaire*. Unpublished manual. University of Minnesota; 1982.
27. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8:141–4.
28. Beurskens AJ, de Vet HC, Köke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996;65:71–6.
29. Gommans IHB, Koes BW, van Tulder MW. Validiteit en responsiviteit van de Nederlandsetalige Roland Disability Questionnaire. *Neder Tijdschr Fysiother* 1997;107:28–33.
30. Miller RP, Kori SH, Todd DD. The Tampa Scale. Tampa, FL; 1991.
31. Vlaeyen JW, Kole Snijders AM, Boeren RG, et al. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain* 1995;62:363–72.
32. Sullivan MJ, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. *Psychol Assess* 1995;7:524–32.
33. Carlsson AM. Assessment of chronic pain. I. aspects of the reliability and validity of the visual analogue scale. *Pain* 1983;16:87–101.
34. Revill SI, Robinson JO, Rosen M, et al. The reliability of a linear analogue for evaluating pain. *Anaesthesia* 1976;31:1191–8.
35. Sriwatanakul K, Kelvie W, Lasagna L, et al. Studies with different types of visual analog scales for measurement of pain. *Clin Pharmacol Ther* 1983; 34:234–9.
36. Beurskens AJ, de vet HC, Köke AJ, et al. A patient-specific approach for measuring functional status in low back pain. *J Manipulative Physiol Ther* 1999;22:144–8.
37. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. conceptual framework and item selection. *Med Care* 1992;30: 473–83.
38. van der Zee K, Sanderman R, Heyink J. De psychometrische kwaliteiten van de MOS 36-item Short Form Health Survey (SF-36) in een Nederlandse populatie. *T Soc Gezondheidsz* 1993;71:183–91.
39. Bø K, Hilde G, Storheim K. Intra- and interobserver reproducibility of Cybex EDI-320 measuring spinal mobility. *Scand J Med Sci Sports* 1997;7:140–3.
40. Chiarello CM, Savidge R. Interrater reliability of the cybex EDI-320 and

fluid goniometer in normals and patients with low back pain. *Arch Phys Med Rehabil* 1993;74:32–7.
41. Koes B, van Mameren H, Bouter L, et al. De reproduceerbaarheid van metingen aan de wervelkolom met de hoekmeter EDI-320. *Neder Tijdschr Fysiother* 1989;99:207–11.
42. van Damme S, Crombez G, Vlaeyen JW, et al. [The Pain Catastrophizing Scale: psychometric characteristics and norms]. *Gedragstherpie* 2000;33: 211–22.
43. Klaber Moffet J, Torgerson D, Bell-Seyer S, et al. Randomized controlled trial of exercise for low back pain: clinical outcome, costs, and preferences. *Br Med J* 1999;319:279–83.
44. Mannion AF, Muntener M, Taimela S, et al. A randomized clinical trial of three active therapies for chronic low back pain. *Spine* 1999;24:2435–48.
45. Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. *Spine* 1987;12(suppl):S1–S59.
46. Pocock SJ. *Clinical Trials. A Practical Approach*. Chichester, NY: John Wiley; 1991.